Delivery 1

Daniel Salgado Rojo

Data Visualization and Modelling. Course 2017/2018.

6) *Dr. Ann Nicholson spends 60 % of her work time in her office. The rest of her work time is spent elsewhere. When Ann is in her office, half the time her light is off (when she is trying to hide from students and get research done). When she is not in her office, she leaves her light on only 5 % of the time. 80 % of the time she is in her office, Ann is logged onto the computer. Because she sometimes logs onto the computer from home, 10 % of the time she is not in her office, she is still logged onto the computer.*

(a) Construct a Bayesian network to represent the scenario just described.

Solution. The first thing we do is to define the random variables that model the situation described above. In particular, we consider three boolean random variables X_1, X_2 and X_3, with values in the set $\{T, F\}$, defined as follows:

$$X_1 = \begin{cases} \text{Ann is at her office, (T)} \\ \text{Ann is not at her office (F)} \end{cases}$$

$$X_2 = \begin{cases} \text{The light is on, (T)} \\ \text{The light is not on (F)} \end{cases}$$

$$X_3 = \begin{cases} \text{Ann is logged onto the computer, (T)} \\ \text{Ann is not logged onto the computer, (F)} \end{cases}$$

From the problem statement we can extract the following values for different ("observed") probabilities of X_2 and X_3 given X_1:

$$P(X_1 = T) = 0.6, \qquad P(X_1 = F) = 0.4 \tag{1}$$
$$P(X_2 = F/X_1 = T) = 0.5, \qquad P(X_2 = T/X_1 = T) = 0.5 \tag{2}$$
$$P(X_2 = T/X_1 = F) = 0.05, \qquad P(X_2 = F/X_1 = F) = 0.95 \tag{3}$$
$$P(X_3 = T/X_1 = T) = 0.8, \qquad P(X_3 = F/X_1 = T) = 0.2 \tag{4}$$
$$P(X_3 = T/X_1 = F) = 0.1, \qquad P(X_3 = F/X_1 = F) = 0.9 \tag{5}$$

In order to build a Bayesian Network these known probabilities lead to the following DAG Γ (Figure 1):

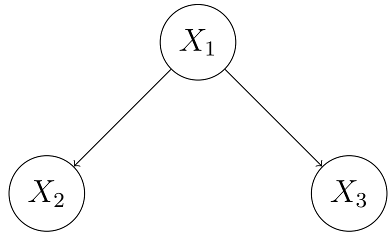

Figure 1: DAG of our BN model.

which we formally define as $\Gamma = (V, E)$ with $V = \{X_1, X_2, X_3\}$ and $E = \{(X_1, X_2), (X_1, X_3)\}$.

We are going to apply the following theorem:

Theorem 1. *(Theorem 2 from the slides) Let Γ be a DAG in which each node is a discrete random variable, and let a conditional distribution of each node given the values of its parents in Γ be specified. Then, the product of these conditional distributions yields a joint probability distribution P of the variables, and (Γ, P) satisfies the Markov condition. In particular, it is a Bayesian Network.*

The probabilities in (1) gives the probability of X_1 given the values of its parents ($\emptyset$, so are the usual probabilities) . The probabilities in (2) and (3) give the probabilities of X_2 given the values of its parents (which is X_1), and the same applies for X_3 with the probabilities in (4) and (5).

Now if we define a joint probability P for (X_1, X_2, X_3) as:

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = P(X_3 = x_3/X_1 = x_1)P(X_2 = x_2/X_1 = x_1)P(X_1 = x_1) \tag{6}$$

for all $x_1, x_2, x_3 \in \{T, F\}$ ("true" or "false), then due to Theorem (1), we have that (Γ, P) is a Bayesian Network with which we model the situation described in the problem statement.

◀

(b) Suppose a student checks Dr. Nicholson's login status and sees that she is logged on. What effect does this have on the student's belief that Dr. Nicholson's light is on?

Solution. The student has observed the event $X_3 = T$. We want to compute the probability that the light is on, i.e. that $X_2 = T$, given that $X_3 = T$: $P(X_2 = T/X_3 = T)$.

Using basic properties of the conditional probabilities we have that:

$$\begin{aligned} P(X_2 = T/X_3 = T) &= \frac{P(X_2 = T, X_3 = T)}{P(X_3 = T)} = \frac{P(X_1 \in \{T, F\}, X_2 = T, X_3 = T)}{P(X_3 = T)} \\ &= \frac{P(X_1 = T, X_2 = T, X_3 = T) + P(X_1 = F, X_2 = T, X_3 = T)}{P(X_3 = T)} \end{aligned}$$

The law of total probability gives:

$$P(X_3 = T) = P(X_3 = T/X_1 = T)P(X_1 = T) + P(X_3 = T/X_1 = F)P(X_1 = F) \quad (7)$$

The definition of our joint probability P as the product of conditional probabilities of the nodes given the values of its parents gives:

$$P(X_1 = T, X_2 = T, X_3 = T) = P(X_3 = T/X_1 = T)P(X_2 = T/X_1 = T)P(X_1 = T) \quad (8)$$

$$P(X_1 = F, X_2 = T, X_3 = T) = P(X_3 = T/X_1 = F)P(X_2 = T/X_1 = F)P(X_1 = F) \quad (9)$$

Now substituting the probability values from equations (1),..., (5) into (7), (8) and (9) we have that

$$\begin{aligned} P(X_3 = T) &= 0.8 \times 0.6 + 0.1 \times 0.4 = 0.52 \\ P(X_1 = T, X_2 = T, X_3 = T) &= 0.8 \times 0.5 \times 0.6 = 0.24 \\ P(X_1 = F, X_2 = T, X_3 = T) &= 0.1 \times 0.05 \times 0.4 = 0.002 \end{aligned}$$

So that finally,

$$\boxed{P(X_2 = T/X_3 = T) = \frac{0.24 + 0.002}{0.52} \approx 0.4654}$$

The fact that the student knows that the professor is logged on leads to the conclusion that the light is on with 46.54% probability. ◀

(c) Is the evidence that Dr. Nicholson's is logged on in favour of or against her light is on?

Solution. To see if the fact of knowing that the professor is logged on $(X_3 = T)$ is in favour or against the fact that her light is on $(X_2 = T)$, is mathematically equivalent to compare $P(X_2 = T/X_3 = T)$ with the prior probability of the light being on, i.e., $P(X_2 = T)$.

To compute $P(X_2) = T$ we use the same probability property we used to compute $P(X_3 = T)$ (which uses now the fact that $P(\{X_1 = T\} \cup \{X_1 = F\}) = 1$ and that this two events are independent):

$$\begin{aligned} P(X_2 = T) &= P(X_2 = T/X_1 = T)P(X_1 = T) + P(X_2 = T/X_1 = F)P(X_1 = F) \\ &= 0.5 \times 0.6 + 0.05 \times 0.4 = 0.32 \end{aligned}$$

thus,

$$\boxed{P(X_2 = T) = 0.32}$$

Since $P(X_2 = T) = 0.32 < 0.4654 = P(X_2 = T/X_3 = T)$, this implies that the fact of knowing that the professor is logged on is in favour of the fact that the light is on (because $X_3 = T$ has increased the probability). ◀